

# SPECIFICATION

Electronic Version 1.2.8

Stylesheet Version 1.0

## [System and Method for Determining User Identity Fraud Using Similarity Searching]

### Cross Reference To Related Applications

#### Referenced-applications

This application claims the benefit of U.S. Provisional Application 60/201073, filed April 26, 2000.

### Background of Invention

- [0001] The current invention relates to identifying occurrences of User Identity Fraud. More specifically, the invention relates to identifying User Identity Fraud by searching for similarities across databases and determining the validity of identity attributes submitted by new users that are attempting to gain access to computer systems.
- [0002] A user or potential user of a computer system commits User Identity Fraud, when the user attempts to gain access to the computer system by knowingly misrepresenting their identifying attributes. Due to past activities, users may be prohibited from accessing information on certain computer systems. These users may attempt to circumvent such restrictions by altering their identity information when attempting to establish new accounts. Consequently, operators of computer systems have been in need of a means to identify these users, by searching across multiple databases for commonalities or similarities in the identity attributes provided by a user when creating a new account. Hence, a system and method have been developed to verify the identities of users who are attempting to establish new accounts, by performing a similarity search for new user identity

attributes across multiple databases. By employing the current invention, operators can better understand who is actually granted access to their computer systems.

## Summary of Invention

- [0003] The current invention provides a method for verifying the identity of a new-user of a computer system, in which at least one identity attribute is received from the new-user and similarity searched against at least one database of denied-user identity attributes. The identity attributes may be received from the new user via Internet web sites, relational databases, data entry systems, and hierarchical databases. The identity attributes received from the new user and stored in the denied-user database may comprise user profiles.
- [0004] A similarity search result is received, which may comprise the results and at least one hierarchical document stored in the at least one database of denied-user identity attributes. A positive or negative match is determined between the at least one new-user identity attribute and the denied-user identity attributes. This determination may comprise comparing the similarity search result to a first match tolerance level, wherein a positive match comprises a match, between the at least one new-user identity attribute and at least one denied-user identity attribute, that meets or exceeds the first match tolerance level; and a negative match comprises a match that does not meet or exceed the first match tolerance level.
- [0005] The new-user is allowed to access the computer system, where a negative match has been determined, and the new-user is denied access to the computer system, where a positive match has been determined. Where a positive match is determined, it may be verified according to a secondary review process. This may include comparing the similarity search result to a second match tolerance level, allowing the new-user to access the computer system where the positive match does not meet or exceed the second match tolerance level, and ultimately denying the new-user access to the computer system, where the positive match meets or exceeds the second match tolerance level.
- [0006] Once a positive or negative match has been determined, the method may then

include the steps of adding the new-user identity to at least one database of valid user identities, where a negative match has been determined, and adding the new-user identity attributes to the at least one database of denied-user identity attributes, where a positive match has been determined.

[0007] The present invention is also directed to a software program embodied on a computer-readable medium incorporating the invented method.

[0008] The present invention is also directed to a system for verifying the identity of a new-user of a computer system, comprising a means for receiving at least one identity attribute from the new-user. The system also comprises at least one database for storing denied-user identity attributes and at least one database for storing valid user identities. The system also comprises a means for similarity searching the at least one identity attribute against the at least one database of denied-user attributes and a means for determining a positive or negative match between the at least one new-user attribute and the at least one database of denied-user identity attributes. The system further comprises a means for allowing the new-user to access the computer system, where a negative match has been determined and a means for denying the new-user access to the computer system, where a positive match has been determined. The system may also comprise a means for adding the new-user identity to the at least one database for storing valid user identities, where a negative match has been determined, and a means for adding the at least one new-user identity attribute to the at least one database of denied-user attributes, where a negative match has been determined.

[0009] The similarity search technique used in the present invention may be any similarity search technique that yields a similarity search result. For example, the similarity search technique used may be that described in U.S. Patent No. 5,666,442 issued September 22, 1999 to Wheeler, which is incorporated by reference herein. The similarity search technique used may also be that described in U.S. Patent Application No. 09/401,101 by Wheeler, et al., filed on September 22, 1999, which is also incorporated by reference herein. Other similarity search techniques may be utilized, as well.

## Brief Description of Drawings

- [0010] FIG. 1 is a block diagram illustrating an overview of the workflow and architecture used to determine user identity fraud via similarity searching, in accordance with the present invention.
- [0011] FIG. 2 is a flow diagram illustrating steps of a method for determining user identity fraud via similarity searching, in accordance with the present invention.
- [0012] FIG. 3 is a block diagram illustrating architecture of a system for determining user identity fraud via similarity searching, in accordance with the present invention.

## Detailed Description

- [0013] Referring to the drawings in detail, the current invention is directed to a system and method for using similarity searching to determine the validity of new user identities. FIG. 1 illustrates an overview 100 of the current invention. The invention uses profile searching to search data from one data feed against another database, in order to find similarities between the two sets of data. A new user profile data feed 101 is transmitted to a similarity search engine (SSE) batch search component 103. The data feed 101 can come from many different sources where data can be acquired. The source of the data feed 101 may comprise, for example, Internet web sites, relational databases, data entry systems, and hierarchical databases.
- [0014] The new user profile data feed 101 is used as an anchor for search component 103 to perform a similarity search against hierarchical documents in a denied users profile database 102. The search component 103 formulates the new user profile data feed 101 into a search request that will be used against the denied user profile database 102. Once the search component 103 completes its search of the denied user profile database 102, a similarity search result set is returned.
- [0015] The similarity search result set is examined by a match component 104 that determines whether a profile match exists. The match component 104 uses pre-defined match tolerance levels to determine whether a profile match exists. The tolerance levels may be set by an operator of the computer system that is served

by the current invention. The results from a similarity profile search fall into two categories: a positive profile match, which meets the specified tolerance level; or a negative profile match, which does not meet the specified tolerance level.

Depending on the category, the results of the search are forwarded to a corresponding component in order to follow a programmed action.

[0016] If the similarity search result set meets the specified profile match tolerance, it is forwarded to the positive response component 105. The positive response component 105 performs predefined actions in response to a positive profile match. For example, if a strong similarity match is made from a new user data feed to a denied user database, a notification can be forwarded to an investigator or to another component for further review.

[0017] If the similarity result search set does not meet the specified tolerance level, then the new user profile is forwarded to the negative response component 106. The negative response component 106 is responsible for executing pre-defined actions for similarity profile results sets that do not meet the specified tolerance level. This could include forwarding the new user profile to a component that stores valid user accounts.

[0018] The denied user profile database 102 can be updated through manual updates, automatically by a profile match triggering system, or periodically through a re-import and re-indexing of the original source profile database.

[0019] FIG. 2 illustrates a workflow for attempting to identify an occurrence of user identity fraud, in accordance with one embodiment of the current invention. The process begins with creating a new user account, in accordance with step 201. The creation of the account may be requested by the new user and may include the transmission of identity attributes to and from Internet web sites, relational databases, data entry systems, and hierarchical databases.

[0020] In accordance with step 202, a similarity search is performed on the new user account attributes, against a denied user profile database. A similarity search result set is returned. In accordance with step 203, the similarity search result set is

checked to see if a profile match exists. The determination of a profile match may include determining whether the results meet a pre-specified match tolerance. The tolerance may be set by a person or persons employing the current invention to identify user identity fraud.

[0021] If the profile does not meet the match tolerance used for the determination in step 203, then the new user account is accepted, in accordance with step 209. The new user account is added to a valid account database, in accordance with step 210, and the new user is allowed all privileges of the account for which the new user applied.

[0022] If the profile meets the match tolerance level used in the determination of step 203, then the similarity search result set for the new user profile is forwarded to a user review database, in accordance with step 204. The user review database stores new user profiles that have matched profiles contained in the denied user database. The new user profile match in the user review database is then re-verified, in accordance with step 205. The validation may proceed according to a pre-defined secondary review process. The secondary review process may again entail determining whether the profile meets a pre-defined match tolerance. The tolerance may be more, less, or equally restrictive, compared to that used for the profile match determination in step 203. If the profile does not meet the specified match tolerance level for the validation in step 206, the new user profile is ultimately accepted, in accordance with step 209. The new user account is then added to a valid account database, in accordance with step 210, and the new user is allowed all privileges of the account for which the new user applied.

[0023] If the new user profile meets the specified match tolerance level used for the validation in step 206, then it is determined that the new user profile is in fact similar to an existing user profile in the denied user database. The new user profile is then denied, in accordance with step 207. Upon denial, the new user profile is added to the denied users database, in accordance with step 208. Thus, the new similarity profile will be included in subsequent searches and determinations of attempts at user identity fraud.

[0024] FIG. 3 illustrates a system architecture for identifying user identity fraud involving similarity searching a new user profile against a set of denied user profiles, in accordance with one embodiment of the current invention. When new user information 301 is provided to the system, it is submitted to the similarity search engine (SSE) batch search component 302. The search component 302 uses the new user information 301 as an anchor for the similarity search criteria that the search component 302 attempts to identify in the denied user database. The search component 302 formulates a similarity search request 303 from the new user information 301 by taking each tagged piece of information contained in the new user profile and adding it to a query with the same tag. The search component 302 then transmits the search request 303 to the similarity search engine server 305.

[0025] The similarity search engine server 305 contains several components. The search request 303 is received into the Gateway component of the similarity search engine server 305. The Gateway component processes the search request 303 and issues search commands to the similarity search engine (SSE). The SSE performs a similarity search across the data bands 306 of the denied users profile database. When results have returned from the SSE, the Gateway then sends requests to pull hierarchical documents from the denied users profile database 304, which is accessed through the File Storage System (FSS). The Gateway then sends search results 307 from the similarity search engine server 305 back to the SSE batch search component 302.

[0026] The similarity search results 307 may comprise the results from the search of the new user profile against the denied users database. The search results 307 may also comprise hierarchical documents retrieved from the denied users profile database 304, which contain the user information for the matching profiles.

[0027] When the similarity search results 307 are returned to the cross- database search component 302, the search component 302 determines whether there is a profile match. This determination may include the use of a pre-specified match tolerance level, as described previously. If a profile match does not exist, then the

search results are forwarded as accepted new user data 311 , to a valid user database 312 for storage, and the new user is allowed future access to the system. If a profile match exists, the search results are forwarded as a hierarchical similarity search result set 308 , to a secondary new user profile review component 309 , for storage and review.

[0028] The hierarchical similarity search result set 308 is added to a user review database in preparation for a secondary review process that ultimately confirms or denies the validity of the new user. The secondary user review component 309 uses a predefined secondary review process to determine whether the new user profile is in fact similar to the denied user profile, within a match tolerance level that may be pre-defined by a system operator. If the user profile meets the match tolerance level, then the new user account is denied, and the denied new user data 310 is then added to the denied user profile database 304 via the Gateway of the similarity search engine server 305 .

[0029] If the new user profile does not meet the match tolerance level used during the secondary review process, then the new user profile is forwarded as accepted new user data 311 to a valid user database 312 for storage, and the new user is allowed future access to the system.

[0030] The current invention is also directed to a software program embodied on computer-readable media, incorporating the method of the current invention.

[0031] Using the foregoing, the invention may be implemented using standard programming or engineering techniques including computer programming software, firmware, hardware or any combination or subset thereof. Any such resulting program, having a computer readable program code means, may be embodied or provided within one or more computer readable or usable media, thereby making a computer program product, i. e. an article of manufacture, according to the invention. The computer readable media may be, for instance a fixed (hard) drive, disk, diskette, optical disk, magnetic tape, semiconductor memory such as read-only memory (ROM), or any transmitting/receiving medium such as the Internet or other communication network or link. The article of



manufacture containing the computer programming code may be made and/or used by executing the code directly from one medium, by copying the code from one medium to another medium, or by transmitting the code over a network.

[0032] An apparatus for making, using or selling the invention may be one or more processing systems including, but not limited to, a central processing unit (CPU), memory, storage devices, communication links, communication devices, server, I/O devices, or any sub-components or individual parts of one or more processing systems, including software, firmware, hardware or any combination or subset thereof, which embody the invention as set forth in the claims.

[0033] User input may be received from the keyboard, mouse, pen, voice, touch screen, or any other means by which a human can input data to a computer, including through other programs such as application programs.

[0034] Although the present invention has been described in detail with reference to certain embodiments, it should be apparent that modifications and adaptations to those embodiments may occur to persons skilled in the art without departing from the spirit and scope of the present invention as set forth in the following claims.